Genomic Prediction and QTL Mapping Using Bayesian Methods

A.S. Leaflet R2647

Xiaochen Sun, research assistant;
David Habier, research scientist, Pioneer Hi-Bred Int.;
Rohan Fernando, professor;
Dorian Garrick, professor;
Jack Dekkers, professor, Department of Animal Science,
Iowa State University

Summary and Implications

Several genomic selection methods were applied to a data set that was simulated for the 2010 QTLMAS workshop to predict the genomic breeding values (GEBV) of the offspring generation and to map the QTL. The GEBV had an accuracy of 0.894 with very small bias. QTL were detected based on the variance of 10 SNP windows. Using a threshold chosen for a 10% chromosome-wise type-I error rate, most of the large QTL were successfully detected with few false positives. Results for both prediction of breeding values and detection of QTL were among the best among all analyses of this data set by groups across the globe. Genomic selection method BayesCπ was identified to be appropriate for the 2010 QTLMAS dataset and also applicable to real cases with similar settings.

Introduction

The availability of high density SNP genotypes across the whole genome has enabled more accurate prediction of breeding values than conventional pedigree-based methods, as well as mapping QTL across the genome. The large number of available SNPs, however, raises the problem that the number of SNP effects to be estimated is usually much greater than the number of phenotypic records. Further, one QTL might be in linkage disequilibrium (LD) with multiple SNPs, which adds noise to the signals for QTL mapping. Bayesian model averaging methods address these problems by fitting all SNPs simultaneously and by shrinking small effects toward zero. This increases the accuracy of detecting QTL, in particular when grouping the effects of neighboring SNPs using the variance of genomic EBV of SNP windows. To enable comparison of methods used by different research groups across the globe, a QTLMAS workshop is held each year, which includes analysis of a simulated data set. The objective of this study was to identify the accuracy of Bayesian methods in predicting EBV and for detecting QTL for the QTLMAS 2010 workshop data set.

Materials and Methods

The 2010 QTLMAS workshop provided simulated data on 10,031 SNP genotypes across a genome of 5 Morgans on 3,226 animals in 5 pedigreed generations. The first four generations had phenotypic records for a quantitative trait and the objective was to predict the breeding values of individuals in generation 5 and to map the QTL. Several genomic selection methods, as implemented in the GenSel software developed at Iowa State University (Fernando and Garrick) were used to analyze the phenotypes and genotypes of the first three generations. The best method was identified by comparing predicted GEBV from each method with phenotypes of individuals in the fourth generation. The effect of fitting polygenic effects into the model was also investigated. Using the best model (BayesCπ), the GEBV of the fifth generation were predicted using the SNP effects estimated from 2,326 animals in the first four generations and sent to the workshop organizers for comparison to the true simulated genotypic values of these animals.

For QTL mapping, SNP effects were estimated from analysis of data from the first four generations with method BayesCπ. QTL regions were identified based on the variance of GEBV of 10 consecutive SNPs. To declare significance, a 10% chromosome-wise threshold was derived by simulation.

Results and Discussion

For the 2010 QTLMAS data, method BayesCπ had higher accuracy and less bias than the other methods that were implemented (see Table 1). Inclusion of a polygenic effect had limited impact. The correlation between GEBV and true genotypic values in the fifth generation was 0.894. On average 124 SNPs were fitted in the model and this was sufficient to explain most of the genetic variance. The use of window variances allowed detection of 16 of the 30 QTLs that were used to simulate the data, with only two false positives. QTL with small effects and one imprinted QTL were not detected. The model that was implemented only captures additive effects of QTL, and advanced methods accounting for higher-order interactions and that are efficient for detecting small QTL remain to be developed. Nevertheless, our analyses resulted in among the highest accuracy of GEBV and identification of the most QTL with fewest false positives compared to other groups that analyzed this same data.

Acknowledgments

Table 1. Prediction accuracy of GEBV and regression coefficients of phenotype on GEBV in the 4th generation using different methods of analysis, as implemented in the GenSel program.

Method	Pedigree BLUP	Genomic BLUP		BayesB, 0.75[1]		BayesB, 0.99		BayesCπ	
		No Poly	Poly[2]	No Poly	Poly	No Poly	Poly	No Poly	Poly
Accuracy	0.545	0.746	0.737	0.781	0.778	0.793	0.790	0.796	0.796
Regression	1.156	1.006	0.961	1.018	0.984	1.031	0.981	1.011	0.989

[1] $\pi = 0.75$, so on average 75% of SNPs were assumed to have ignorable small effects and were not fitted into the model.

[2] No Poly = model without polygenic effects; Poly = model with polygenic effects.